September 11, 2009

ON IMPORTANCE SAMPLING WITH MIXTURES FOR RANDOM WALKS WITH HEAVY TAILS

HENRIK HULT AND JENS SVENSSON

Abstract. Importance sampling algorithms for heavy-tailed random walks are considered. Using a specification with algorithms based on mixtures of the original distribution with some other distribution, sufficient conditions for obtaining bounded relative error are presented. It is proved that mixture algorithms of this kind can achieve asymptotically optimal relative error. Some examples of mixture algorithms are presented, including mixture algorithms using a scaling of the original distribution, and the bounds of the relative errors are calculated. The algorithms are evaluated numerically in a simple setting.

1. Introduction

Tail probabilities appear naturally in many applications of probability theory, and often analytical evaluation is not possible. For many applications, Monte Carlo simulation can be an effective alternative. For rare events, however, standard Monte Carlo simulation is very computationally inefficient, and some form of variance reduction method is necessary. One such alternative that has been extensively applied to both light- and heavy-tailed distributions is importance sampling. In this paper we focus on importance sampling algorithms for computing the probability

$$p_b = P(S_n > b),$$

of a high threshold b, for a random walk $S_n = X_1 + \cdots + X_n$. The random variables $X_1, \dots, X_n$ are independent and identically distributed with distribution function F and density f. It is assumed that the right tail of f is regularly varying at ∞; more precisely there exists an $\alpha > 0$ such that, for each $x > 0$,

$$\lim_{u \to \infty} \frac{f(ux)}{f(u)} = x^{-\alpha-1}.$$

Then it is well known that f has the representation $f(x) = x^{-\alpha-1}L(x)$, $x > 0$, where L is slowly varying. The joint distribution of $(X_1, \dots, X_n)$ is denoted μ_n.

Consider first a computation of p_b using standard Monte Carlo. Then N independent samples $(X_1^1, \dots, X_n^1), \dots, (X_1^N, \dots, X_n^N)$ are generated from μ_n and p_b is estimated using the sample frequency

$$\hat{p}_b^{MC} = \frac{1}{N} \sum_{i=1}^{N} I\{S_n^i > b\},$$

where $S_n^i = X_1^i + \cdots + X_n^i$. For large b, the event $\{S_n > b\}$ is rare and few of the indicator variables $I\{S_n^i > b\}$ will be 1. This leads to rather inefficient estimation.

2000 *Mathematics Subject Classification.* Primary: 65C05; Secondary: 60G50.

Key words and phrases. Monte Carlo simulation; rare events; importance sampling; heavy tails.

To see this, consider for instance the standard deviation of $\hat{p}_b^{MC}$. An elementary calculation shows

$$\mathrm{Stdev}(\hat{p}_b) = \frac{1}{\sqrt{N}}\sqrt{p_b(1-p_b)}.$$

When p_b is small this is roughly $\sqrt{p_b/N}$. Hence, it would require $N \approx 1/p_b$ samples to have the standard deviation of size comparable to the quantity p_b we are estimating. When p_b is small this can be very large.

Importance sampling provides a way to possibly reduce the computational cost without sacrificing precision, or equivalently to improve precision without increasing the computational cost. The basic idea of importance sampling to generate samples $(X_1^1,\dots,X_n^1),\dots,(X_1^N,\dots,X_n^N)$ independently from a sampling measure ν_n^b instead of μ_n. It is assumed that μ_n is absolutely continuous with respect to ν_n^b, written $\mu_n \ll \nu_n^b$ so that the Radon-Nikodym derivative $\frac{d\mu_n}{d\nu_n^b}$ exists. An unbiased estimate of p_b is constructed as

$$\hat{p}_b = \frac{1}{N}\sum_{i=1}^{N} \frac{d\mu_n}{d\nu_n^b}(X_1,\dots,X_n) I\{S_n > b\}.$$

The goal is to choose ν_n^b so more samples are drawn from regions that are "important" to the event $\{S_n > b\}$. Then the event becomes less rare under ν_n^b, which reduces variance. However, ν_n^b must be chosen carefully so that the Radon-Nikodym weights $\frac{d\mu_n}{d\nu_n^b}(X_1,\dots,X_n)$ do not cause variance to increase. A relevant quantity for deciding if a sampling measure ν_n^b is appropriate or not is the relative error

$$RE(\hat{p}_b) = \frac{\sqrt{\mathrm{Var}(\hat{p}_b)}}{E\hat{p}_b} = \frac{1}{\sqrt{N}}\sqrt{\frac{E\hat{p}_b^2 - p_b^2}{p_b^2}} = \frac{1}{\sqrt{N}}\sqrt{\frac{E\hat{p}_b^2}{p_b^2} - 1}.$$

By Jensen's inequality we always have $E\hat{p}_b^2 \geq p_b^2$.

To quantify the efficiency of the sampling measure it is convenient to study the asymptotics of the relative error as $b \to \infty$. This amounts to studying the asymptotics of normalized second moment $\lim_{b\to\infty} E\hat{p}_b^2/p_b^2$. We say that a sampling distribution ν_n^b has *logarithmically efficient relative error* if, for some $\varepsilon > 0$,

$$\limsup_{b\to\infty} \frac{E\hat{p}_b^2}{p_b^{2-\varepsilon}} < \infty,$$

it has *bounded relative error* if

$$\limsup_{b\to\infty} \frac{E\hat{p}_b^2}{p_b^2} < \infty,$$

and *asymptotically optimal relative error* if

$$\limsup_{b\to\infty} \frac{E\hat{p}_b^2}{p_b^2} = 1.$$

A number of different algorithms have been proposed to simulate tail probabilities of heavy-tailed random walks. Asmussen and Binswanger (1997) study the class of subexponential distributions, i.e. distributions for which

$$\lim_{b\to\infty} \frac{P(S_n > b)}{nP(X_1 > b)} = 1,$$

and use that as $b \to \infty$, all the variance of the sum comes from the largest summand. By removing the largest term $X_{(n)}$ in each sample and calculating the probability using the remaining $n-1$ terms, they obtain a logarithmically efficient conditional Monte Carlo estimator in the sub-class of distributions with regularly varying tails. Here $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ is the ordered sample. Specifically, with the sample $X_1, \ldots, X_n$, the estimator is

$$P(S_n > b | X_{(1)}, \ldots, X_{(n-1)}) = \frac{\overline{F}(X_{(n-1)} \vee (b - S_{(n-1)}))}{\overline{F}(X_{(n-1)})},$$

where $S_{(n-1)} = X_{(1)} + \ldots X_{(n-1)}$ is the sum of the $n-1$ largest of the n terms in the sample and $a \vee b = \max(a, b)$.

Asmussen and Kroese (2006) propose a similar idea, using the conditioning

$$nP(S_n > b, M_n = X_n | X_1, \ldots, X_{n-1}) = \overline{F}(M_{n-1} \vee (b - S_{(n-1)})),$$

where $M_n = \max(X_1, \ldots, X_n)$ to obtain an estimator with bounded relative error for distributions with regularly varying tails.

Juneja and Shahabuddin (2002) introduce an importance sampling algorithm with similar structure to exponential twisting in the light-tailed case. Their so-called hazard rate twisting of the original distribution is given by

$$dF_\theta(x) = \frac{e^{\theta\Lambda(x)} dF(x)}{\int_0^\infty e^{\theta\Lambda(x)} dF(x)}, \tag{1.1}$$

where $0 < \theta < 1$ and $\Lambda(x) = -\log \overline{F}(x)$ is the hazard rate. For distributions with regularly varying tails, this is equivalent to changing the tail index of the distribution.

In the case of importance sampling, Bassamboo et al. (2007) show that to obtain efficient sampling distributions for heavy-tailed random walks, one must consider state-dependent changes of measure. Simply changing the parameters in the original distribution cannot lead to an estimator with bounded relative error.

The first algorithm of this type, for heavy-tailed random walks, was proposed by Dupuis et al. (2007). There the large values are sampled from the conditional distribution where one has to condition on exceeding a level just below the remaining distance to b. The authors prove that their proposed algorithm has close to asymptotically optimal relative error. Blanchet and Li (2008) present a state-dependent algorithm that uses Markov chain description of the random walk under the sampling measure to obtain bounded relative error for the class of subexponential distributions. Blanchet and Liu (2008) construct a mixture algorithm with bounded relative error for the large deviation probability $P(S_n > b)$ where $b > b_0 n^{1/2+\epsilon}$.

In this paper we take a more general look at mixture algorithms of the same type as Dupuis et al. (2007). The underlying idea is to construct a dynamic change of measure such that the trajectories of $X_1, \ldots, X_n$ leading to $S_n > b$ is similar to the most likely trajectories conditional on $S_n > b$. In the heavy-tailed case, the most likely trajectories are such that one of the X_i's is large and the others are "average". Mixtures arise quite naturally as sampling distributions for producing such trajectories; with some probability p_i sample from the original density f and with probability $q_i = 1 - p_i$ sample from a density where it is likely to get a large value. We provide sufficient conditions for bounded relative error and provide a couple of new examples that are very easy to implement. We also show that, with some additional work, one can construct mixture algorithms with asymoptotically optimal relative error.

The paper is organized as follows. In Section 2 we present a general importance sampling algorithm based on mixtures and provide several examples. In Section 3 we provide sufficient condition for the mixture algorithm to have bounded relative error. In Section 4 we provide detailed analysis of specific mixture algorithms. The concluding Section 5 provides a proof that it is possible to obtain asymptotically optimal relative error.

2. Dynamic mixture algorithms

In this section we describe a general importance sampling algorithm based on mixtures, called the dynamic mixture algorithm, and provide several examples.

The dynamic mixture algorithm for computing $p_b = P(S_n > b)$ proceeds as follows. Each replication of $(X_1, \dots, X_n)$ is generated dynamically and the distribution for sampling X_i depend on the current state $S_{i-1} = X_1 + \cdots + X_{i-1}$ of the random walk. At the ith step it may be that S_{i-1} already exceeds the threshold b. Then X_i is sampled from the original density f. Otherwise, if $S_{i-1} \leq b$, a biased coin is tossed with probability p_i for "heads" and $q_i = 1 - p_i$ for "tails". If it comes up "heads" X_i is generated from the usual density f, but if it comes up "tails", X_i is generated from another density $g_i(x \mid S_{i-1})$. The density $g_i(x \mid S_{i-1})$ depends on the current generation i of the algorithm and on the current position S_{i-1}. The idea is to choose $g_i(x \mid S_{i-1})$ s.t. sampling from $g_i(x \mid S_{i-1})$ is likely to result in a large variable. However, $g_i(x \mid S_{i-1})$ must be chosen with some care to control the Radon-Nikodym weights $\frac{d\mu_n}{d\nu_n^b}(X_1, \dots, X_n)$ and thereby the relative error. In the last generation, if $S_{n-1} \leq b$, X_n is sampled from a density $g_n(x \mid S_{n-1})$ and if $S_{n-1} > b$ it is sampled from the original f. In contrast to the previous steps g_n is not necessarily of mixture type. The reason is that it may be advantageous to make sure $X_n > b - S_{n-1}$ in the last step to get $S_n > b$.

A precise description of the dynamic mixture algorithm is presented next.

Algorithm 1. Consider step $i = 1, \dots, n$, where $S_{i-1} = s_{i-1}$. Then X_i is sampled as follows.

- If $s_{i-1} > b$, X_i is sampled from the original density f,
- if $s_{i-1} \leq b$, X_i is sampled from

$$\begin{aligned} &p_i f(\cdot) + q_i g_i(\cdot \mid s_{i-1}), \qquad \text{for } 1 \leq i \leq n-1, \\ &g_n(\cdot \mid s_{n-1}), \qquad \text{for } i = n. \end{aligned}$$

Here $p_i + q_i = 1$ and $p_i \in (0,1)$.

Explicit examples of the dynamic mixture algorithm are obtained by specifying g_i and p_i.

Example 2.1 (Conditional mixture, c.f. Dupuis et al. (2007))**.** The algorithm proposed by Dupuis et al. (2007) takes g_i to be a conditional distribution. For $i = 1, \dots, n-1$ the large values are sampled conditional on being at least a times the remaining distance to b, where $a \in (0,1)$. It is important that $a < 1$. In the last step samples are generated conditional on exceeding b. More precisely,

$$g_i(x \mid s) = \frac{f(x)I\{x > a(b-s)\}}{\overline{F}(a(b-s))}, \quad 1 \le i \le n-1,$$
$$g_n(x \mid s) = \frac{f(x)I\{x > b-s\}}{\overline{F}(b-s)}.$$

In their paper the authors assume that $f = 0$ on $(-\infty, 0)$. That is, all the X_i's are non-negative. This is not an important restriction and we do not impose it here.

A practical limitation of the conditional mixture algorithm is that some distributions do not allow direct sampling from the conditional distribution. If the distribution function F and its inverse $F^{\leftarrow}$ are available, the inversion method suggest sampling X conditional on $X > c$ by taking U to be uniform on $(0,1)$ and set $X = F^{\leftarrow}(1 - U\overline{F}(x))$, see e.g. Asmussen and Glynn (2007). In other cases it might be necessary to use an acceptance-rejection method, but this may be time consuming.

A simple alternative to the conditional mixture is to sample the large variables from a generalized Pareto distribution (GPD) instead. The intuition is that the GPD approximates the conditional distribution well.

Example 2.2 (Generalized Pareto mixture)**.** The GPD mixture algorithm takes g_i to be a generalized Pareto distribution. As in the previous algorithm, for $i = 1, \dots, n-1$, the large values are sampled conditional on being at least a times the remaining distance to b, where $a \in (0,1)$. The last step is slightly different. If the remaining distance is large, the last step is taken from a GPD, otherwise it is taken from the original density. This is because, if $S_{n-1} \le b$, but close to b, the GPD is not necessarily a good approximation of the conditional distribution. To be precise,

$$\begin{aligned} g_i(x \mid s) &= \alpha[a(b-s)]^{\alpha} x^{-\alpha-1} I\{x > a(b-s)\}, \quad 1 \le i \le n-1. \\ g_n(x \mid s) &= \alpha(b-s)^{\alpha} x^{-\alpha-1} I\{x > b-s\} I\{s \le b - b(1-a)^{n-1}\} \\ &\quad + f(x) I\{s > b - b(1-a)^{n-1}\}. \end{aligned}$$

A different way to sample the large variables is to sample from the original density and then scale the outcome by simply multiplying with a large number λb. We call this a scaling mixture algorithm.

Example 2.3 (Scaling mixtures)**.** The scaling mixture algorithm has, with $\lambda > 0$,

$$\begin{aligned} g_i(x \mid s) &= (\lambda b)^{-1} f(x/\lambda b) I\{x > 0\} + f(x) I\{x \le 0\}, \quad i = 1, \dots, n-1, \\ g_n(x \mid s) &= (\lambda b)^{-1} f(x/\lambda b) I\{x > 0, s \le b - b(1-a)^{n-1}\} \\ &\quad + f(x) I\{x \le 0 \text{ or } s > b - b(1-a)^{n-1}\}. \end{aligned}$$

To simplify the analysis we will, in the context of scaling mixtures, always assume that the orginal density f is strictly positive on $(0, \infty)$. If this is not satisfied the situation is more involved because there may be large $x > 0$ such that $f(x) > 0$ but $f(x/\lambda b) = 0$. Then such large x-value cannot be obtained by sampling a small number from f and scale by λb. This may cause the Radon-Nikodym weights to be relatively large, which increase the variance.

There are several variations of the scaling algorithm. For instance, one may scale with something proportional to the remaining distance to b, instead of something proportional to b as described above. Some variations of the scaling algorithm will be treated in more detail in Section 4.3.

3. Asymptotic analysis of the normalized second moment

The efficiency criteria presented in the introduction are all based on the asymptotic properties of the normalized second moment $E\hat{p}_b^2/p_b^2$. We are following the weak convergence approach initiated by Dupuis et al. (2007) to study its asymptotics. By the subexponential property, $p_b^2 \sim n^2\overline{F}(b)^2$, where $a_b \sim c_b$ denotes $\lim_{b\to\infty} a_b/c_b = 1$, the normalized second moment can be written as

$$\frac{E\hat{p}_b^2}{p_b^2} \sim \frac{1}{n^2\overline{F}(b)^2}\int\limits_{s_n>b} \frac{d\mu_n}{d\nu_n^b}(y)\mu_n(dy) = \frac{1}{n^2}\int\limits_{s_n>1} \frac{1}{\overline{F}(b)}\frac{d\mu_n}{d\nu_n^b}(by)m_b(dy), \tag{3.1}$$

where the measure $m_b = \mu_n(b(\,\cdot \cap \{s_n > 1\}))/\overline{F}(b)$. To calculate the limit of this integral we will use the weak convergence of the measure m_b to a measure m and uniform convergence of an upper bound $R_b^*(y) \geq \frac{1}{\overline{F}(b)}\frac{d\mu_n}{d\nu_n^b}(by) =: R_b(y)$ to a bounded continuous function $R(y)$. Then we (?) establish the convergence

$$\limsup_{b\to\infty} \int\limits_{\{s_n>1\}} R_b dm_b \leq \lim_{b\to\infty}\int\limits_{\{s_n>1\}} R_b^* dm_b = \int\limits_{\{s_n>1\}} R dm.$$

To do this it is convenient if the normalized Radon-Nikodym derivative $R_b(y)$ is bounded. This criteria is certainly stronger than necessary but appears to be desirable. It implies the the normalized q-moment is asymptotically bounded for any $q \in (1,\infty)$. Indeed, if R_b^* is bounded and $R^* \to R$ uniformly, then for any $q \in (1,\infty)$

$$\limsup_{b\to\infty}\frac{E\hat{p}_b^q}{p_b^q} = \limsup_{b\to\infty}\frac{1}{n^q}\int\limits_{s_n>1}\Big(\frac{1}{\overline{F}(b)}\frac{d\mu_n}{d\nu_n^b}(by)\Big)^{q-1} m_b(dy) \leq \frac{1}{n^q}\int R^{q-1}dm < \infty.$$

Next we provide sufficient conditions for R_b to be bounded.

Lemma 3.1. *Consider Algorithm 1 with $p_i > 0$ for $1 \leq i \leq n-1$. Suppose there exists $a \in (0,1)$ such that*

$$\liminf_{b\to\infty} \inf_{\substack{s \leq 1-(1-a)^{i-1}\\ y > a(1-s)}} \frac{g_i(by \mid bs)}{f(by)}\overline{F}(b) > 0, \quad 1 \leq i \leq n, \tag{3.2}$$

$$\limsup_{b\to\infty} \sup_{\substack{s \leq 1\\ y > 1-s}} \frac{f(by)}{g_n(by \mid bs)} < \infty. \tag{3.3}$$

Then the scaled Radon-Nikodym derivative $R_b(y) = \frac{1}{\overline{F}(b)}\frac{d\mu}{d\nu_n^b}(by)$ is bounded on $\{y_1 + \cdots + y_n > 1\}$.

Proof. Let $s_n = y_1 + \cdots + y_n$. On $\{s_n > 1\}$ it must hold that $y_i > a(1-s_{i-1})$ for some $i = 1,\dots,n$. Otherwise $s_i \leq 1-(1-a)^i < 1$ for each i.

Take $y \in \{y \in \mathbb{R}^n : s_n > 1\}$ and let $i = \min\{j : y_j > a(1-s_{j-1})\}$. Note that for this i

$$y_i > a(1-s_{i-1}) \geq a(1-a)^{i-1} \geq a(1-a)^n =: a_n > 0.$$

For any y_j, $j \notin \{i, n\}$,

$$\frac{f(by_j)}{p_j f(by_j) + q_j g_j(by_j \mid bs_{j-1})} \leq \frac{1}{p_j}.$$

It follows that, for $1 \leq i \leq n-1$,

$$\begin{aligned}\frac{1}{\overline{F}(b)} \frac{d\mu}{d\nu_n^b}(by) \leq \frac{1}{\overline{F}(b)} \frac{f(by_i) I\{y_i > a(1 - s_{i-1})\}}{p_i f(by_i) + q_i g_i(by_i \mid bs_{i-1})} \times \prod_{j \notin \{i,n\}} \frac{1}{p_j} \\ \times \Big(\frac{f(by_n) I\{y_n > 1 - s_{n-1}\}}{g_n(by_n \mid bs_{n-1})} I\{s_{n-1} \leq 1\} + I\{s_{n-1} > 1\} \Big). \quad (3.4)\end{aligned}$$

The first term can be written as

$$\frac{1}{\overline{F}(b)} \frac{f(by_i) I\{y_i > a(1 - s_{i-1})\}}{p_i f(by_i) + q_i g_i(by_i \mid bs_{i-1})} = \frac{I\{y_i > a(1 - s_{i-1})\}}{p_i \overline{F}(b) + q_i \frac{g_i(by_i \mid bs_{i-1})}{f(by_i)} \overline{F}(b)}.$$

By (3.2) this term in (3.4) is bounded. The second term is bounded because $p_j > 0$ by assumption. The last term is bounded by (3.3).

Similarly for $i = n$,

$$\frac{1}{\overline{F}(b)} \frac{d\mu}{d\nu_n^b}(by) \leq \frac{f(by_n) I\{y_n > a(1 - s_{n-1})\}}{g_n(by_n \mid bs_{n-1}) \overline{F}(b)} \prod_{j=1}^{n-1} \frac{1}{p_j},$$

which is bounded by (3.2). □

Next we present the main result. It provides sufficient conditions for the mixture algorithms to have bounded relative error. This is obtained by showing that the normalized second moment remains bounded.

Theorem 3.2. *Suppose* (3.2) *and* (3.3) *hold for* $a \in (0, 1)$. *Suppose in addition that there exist continuous functions* $h_i : \mathbf{R}^n \to [0, \infty)$ *such that*

$$\frac{f(by_i)}{g_i(by_i \mid bs_{i-1}) \overline{F}(b)} \to h_i(y_i \mid s_{i-1}), \quad (3.5)$$

uniformly on $\{y \in \mathbf{R}^n : s_{i-1} \leq 1 - (1 - a)^{i-1}, y_i > a(1 - s)\}$. *Then,*

$$\lim_{b \to \infty} \frac{E\hat{p}_b^2}{p_b^2} \leq \frac{1}{n^2} \sum_{i=1}^{n} \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \int_1^\infty h_i(y_i \mid 0) \alpha y_i^{-\alpha - 1} dy_i,$$

with the convention that $q_n = 1$.

Proof. First rewrite the normalized second moment as in (3.1):

$$\frac{E\hat{p}_b^2}{p_b^2} \sim \frac{1}{n^2} \int\limits_{s_n > 1} \frac{1}{\overline{F}(b)} \frac{d\mu_n}{d\nu_n^b}(by) m_b(dy) = \frac{1}{n^2} \int R_b(y) m_b(dy).$$

By regular variation of f and independence of $X_1, \ldots, X_n$ the joint distribution μ_n is multivariate regularly varying. In particular the weak convergence $m_b \xrightarrow{w} m$ holds, where m has the representation

$$m(A) = \sum_{i=1}^{n} \int_A I\{y \in \mathbf{R}^n : y_i > 1, y_j = 0, j \neq i\} \alpha y_i^{-\alpha - 1} dy_i. \quad (3.6)$$

This is well known, see e.g. Resnick (1987), Section 5.5. A proof is also given by Dupuis et al. (2007). We see that the measure m puts all its mass on the coordinate axes. That is, on trajectories where one jump is large and the rest are zero.

The next step is to decompose the integral as

$$\int R_b dm_b = \int_A R_b dm_b + \int_{A^c} R_b dm_b, \tag{3.7}$$

where $A = \cup_{i=1}^n A_i$ is a finite union and the A_i's have disjoint closures. We will find A_i such that the second integral converges to 0 and determine an upper bound $R_b^* \geq R_b$ on A_i.

Define the sets A_i to be

$$\begin{aligned} A_i = \{y \in \mathbb{R}^n : y_j \leq a(1 - s_{j-1}) \text{ for } 1 \leq j \leq i-1, y_i > 1 - s_{i-1}, \\ \text{and } s_k > 1, k = i+1, \dots, n\}. \end{aligned}$$

Note that the A_i's have disjoint closure and $m(\partial A_i) = 0$. In particular $m_b(A_i) \to m(A_i)$ for each $i = 1, \dots, n$. Moreover, $m(\cap_{i=1}^n A_i^c) = 0$. Indeed,

$$\begin{aligned} m(\cap_{i=1}^n A_i^c) &= m(\{s_n > 1\} \setminus \cup_i A_i) = m\{s_n > 1\} - \sum_{i=1}^n m(A_i) \\ &= \sum_{i=1}^n \int_{\{s_n > 1\}} I\{y \in \mathbf{R}^n : y_i > 1, y_j = 0, j \neq i\} \alpha y_i^{-\alpha-1} dy_i \\ &\quad - \sum_{i=1}^n \int_{A_i} I\{y \in \mathbf{R}^n : y_i > 1, y_j = 0, j \neq i\} \alpha y_i^{-\alpha-1} dy_i = 0. \end{aligned}$$

By Lemma 3.1, R_b is bounded and since $m_b(A^c) \to m(A^c) = 0$, the second integral in (3.7) converges to 0. For the first integral we construct a function R_b^* that dominates R_b on A and a continuous function R such that $R_b^* \to R$ uniformly on A. Then it follows from weak convergence that

$$\limsup_{b \to \infty} \int_A R_b dm_b \leq \lim_{b \to \infty} \int_A R_b^* dm_b = \int_A R dm < \infty.$$

For $y \in A_i$,

$$R_b(y) \leq \frac{1}{\overline{F}(b)} \frac{f(by_i) I\{y_i > 1 - s_{i-1}\}}{p_i f(by_i) + q_i g_i(by_i \mid bs_{i-1})} \prod_{j=1}^{i-1} \frac{1}{p_j} =: R_b^*(y).$$

To see this, construct a bound as in (3.4) and notice that on A_i, $s_k > 1$ for each $k \geq i$. Then the contribution to the Radon-Nikodym weights from y_k, $k > i$ is equal to 1. By assumption (3.5)

$$\frac{g_i(by_i \mid bs_{i-1})}{f(by_i)} \overline{F}(b) \to \frac{1}{h_i(y_i \mid s_{i-1})},$$

uniformly on A_i. For $y \in A_i$ define $R(y) = h_i(y_i \mid s_{i-1}) \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i}$. Then $R_b^* \to R$ uniformly on A. With the representation (3.6) of the limiting measure m, the upper bound for the normalized second moment can now be calculated as

$$\frac{1}{n^2} \int_A R dm = \frac{1}{n^2} \sum_{i=1}^n \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \int_1^\infty h_i(y_i \mid 0) \alpha y_i^{-\alpha-1} dy_i.$$

□

4. Examples

In this section we provide a detailed analysis of the algorithms presented in Section 2. In particular we verify the conditions of Lemma 3.1 and Theorem 3.2 for these algorithms.

4.1. **The conditional mixture algorithm.**

Recall from Example 2.1 that the conditional mixture algorithm has, with $a \in (0,1)$,

$$\begin{aligned} g_i(x \mid s) &= \frac{f(x) I\{x > a(b-s)\}}{\overline{F}(a(b-s))}, \quad 1 \le i \le n-1, \\ g_n(x \mid s) &= \frac{f(x) I\{x > b-s\}}{\overline{F}(b-s)}. \end{aligned}$$

Then, for $i = 1, \dots, n-1$, the uniform convergence $\overline{F}(bx)/\overline{F}(b) \to x^{-\alpha}$, for $x > x_0 > 0$, implies

$$\frac{g_i(bx \mid bs)}{f(bx)} \overline{F}(b) = \frac{\overline{F}(b)}{\overline{F}(ba(1-s))} I\{x > a(1-s)\} \to a^{\alpha}(1-s)^{\alpha} I\{x > a(1-s)\},$$

uniformly for $s \le 1-(1-a)^{i-1}$, $x > a(1-s)$. Similarly,

$$\frac{g_n(bx \mid bs)}{f(bx)} \overline{F}(b) = \frac{\overline{F}(b)}{\overline{F}(b(1-s))} I\{x > 1-s\} \to (1-s)^{\alpha} I\{x > 1-s\},$$

uniformly on $s \le 1-(1-a)^{n-1}$, $x > 1-s$, and

$$\frac{f(bx)}{g_n(bx \mid bs)} = \overline{F}(b(1-s)) \le 1,$$

on $s \le 1$. It follows that both (3.2) and (3.3) are satisified and hence the normalized Radon-Nikodym derivative is bounded.

By the above calculation (3.5) holds with $h_i(y \mid s) = a^{-\alpha}(1-s)^{-\alpha}$, $1 \le i \le n-1$ and $h_n(y \mid s) = (1-s)^{-\alpha}$. It follows from Theorem 3.2 that

$$\lim_{b \to \infty} \frac{E\hat{p}_b^2}{p_b^2} \le \frac{1}{n^2} \int R(y) dm = \frac{1}{n^2} \Big(\sum_{i=1}^{n-1} \frac{a^{-\alpha}}{q_i} \prod_{j=1}^{i-1} \frac{1}{p_j} + \prod_{j=1}^{n-1} \frac{1}{p_j} \Big). \tag{4.1}$$

The right hand side is minimized at

$$p_i = \frac{(n-i-1)a^{-\alpha/2}+1}{(n-i)a^{-\alpha/2}+1}, \quad q_i = 1-p_i, \tag{4.2}$$

with minimum $n^{-2}[(n-1)a^{-\alpha/2}+1]^2$, and it is possible to show that the limit is equal to the right hand side of (4.1), see Dupuis et al. (2007), Lemma 3.2.1. For each n this can be made arbitrarily close to 1 by choosing a close to 1.

4.2. **Generalized Pareto mixture.**

Recall from Example 2.2 that the GPD mixture algorithm has, with $a \in (0,1)$,

$$\begin{aligned} g_i(x \mid s) &= \alpha[a(b-s)]^{\alpha} x^{-\alpha-1} I\{x > a(b-s)\}, \quad 1 \le i \le n-1. \\ g_n(x \mid s) &= \alpha(b-s)^{\alpha} x^{-\alpha-1} I\{x > b-s\} I\{s \le b-b(1-a)^{n-1}\} \\ &\quad + f(x) I\{s > b-b(1-a)^{n-1}\}. \end{aligned}$$

First we check (3.2) and (3.3). Karamata's theorem implies $\alpha\overline{F}(b) \sim bf(b)$. Then, for any $s < 1$,

$$\begin{aligned}\frac{g_i(bx \mid bs)}{f(bx)}\overline{F}(b) &= \frac{\alpha(bx)^{-\alpha-1}(ba(1-s))^{\alpha}\overline{F}(b)}{f(bx)} \\ &= \frac{\alpha\overline{F}(bx)}{bxf(bx)}\frac{a^{\alpha}(1-s)^{\alpha}\overline{F}(b)}{x^{\alpha}\overline{F}(bx)} \to a^{\alpha}(1-s)^{\alpha}. \qquad (4.3)\end{aligned}$$

uniformly for $x \geq a(1-s)$. In particular (3.2) is satisfied. Since

$$\begin{aligned}\frac{f(by)}{g_n(by \mid bs)} = &\frac{bf(bx)}{\alpha(1-s)^{\alpha}y^{-\alpha-1}}I\{y > 1-s\}I\{s \leq 1-(1-a)^{n-1}\} \\ &+ I\{s > 1-(1-a)^{n-1}\}\end{aligned}$$

is bounded on $s \leq 1$, $y > 1-s$, (3.3) also holds. By Lemma 3.1 the normalized Radon-Nikodym derivative is bounded. By the arguments above (3.5) holds with $h_i(y \mid s) = a^{-\alpha}(1-s)^{-\alpha}$, $1 \leq i \leq n-1$ and $h_n(y \mid s) = (1-s)^{-\alpha}$. It follows by Theorem 3.2 that

$$\lim_{b\to\infty}\frac{E\hat{p}_b^2}{p_b^2} = n^{-2}\Big(\prod_{i=1}^{n-1}\frac{1}{p_i} + a^{-\alpha}\sum_{j=1}^{n-1}\frac{1}{q_j}\prod_{i=1}^{j-1}\frac{1}{p_i}\Big).$$

This is identical to (4.1), so p_i can be chosen according to (4.2) to minimize the relative error.

4.3. **Scaling mixtures.** In the scaling mixture algorithm presented in Example 2.3 the large variables are generated by sampling from the original density and multiplying with a large number. In this section we study some variations of this algorithm. Recall that, in the context of scaling mixtures, always assume that the orginal density f is strictly positive on $(0,\infty)$.

The first scaling mixture algorithm, called *scaling mixture I*, is constructed as follows. Write $f(x) = x^{-\alpha-1}L(x)$ with L slowly varying. Suppose $\inf_{x>x_0} L(x) =: L_* > 0$ for some $x_0 > 0$. The scaling mixture algorithm, with $\lambda > 0$, has

$$\begin{aligned}g_i(x \mid s) &= (\lambda b)^{-1}f(x/\lambda b)I\{x > 0\} + f(x)I\{x \leq 0\}, \quad i = 1,\dots,n-1, \\ g_n(x \mid s) &= (\lambda b)^{-1}f(x/\lambda b)I\{x > 0, s \leq b - b(1-a)^{n-1}\} \\ &\quad + f(x)I\{x \leq 0 \text{ or } s > b - b(1-a)^{n-1}\}.\end{aligned}$$

To generate a sample X from g_i proceed as follows. Generate a candidate X' from f. If $X' \leq 0$ put $X = X'$ and if $X' > 0$, put $X = \lambda b X'$.

Take $a \in (0,1)$, using Karamata's theorem, $\alpha\overline{F}(b) \sim bf(b)$, we have, for $1 \leq i \leq n$, and $s \leq 1-(1-a)^{i-1}$,

$$\begin{aligned}\frac{g_i(bx \mid bs)}{f(bx)}\overline{F}(b) &= \frac{f(\frac{x}{\lambda})}{\lambda bf(bx)}\overline{F}(b)I\{x > 0\} + \overline{F}(b)I\{x \leq 0\} \\ &= \frac{xf(\frac{x}{\lambda})}{\alpha\lambda}\frac{\alpha\overline{F}(bx)}{bxf(bx)}\frac{\overline{F}(b)}{\overline{F}(bx)}I\{x > 0\} + \overline{F}(b)I\{x \leq 0\} \\ &\to \frac{x^{\alpha+1}f(\frac{x}{\lambda})}{\alpha\lambda}\end{aligned}$$

uniformly for $x \geq 1 - s$. Since $x^{\alpha+1} f(x/\lambda) \geq \lambda^{\alpha+1} L_* > 0$, the condition (3.2) holds. Note, however, that (3.2) fails if $L_* = 0$. Since

$$\frac{f(bx)}{g_n(bx \mid bs)} = \frac{\lambda b f(bx)}{f(\frac{x}{\lambda})} I\{x > 0, s \leq 1 - (1-a)^{n-1}\} + I\{x \leq 0 + s > 1 - (1-a)^{n-1}\}$$

is bounded on $s \leq 1$, $x > 1 - s$ condition (3.3) also holds. From the calculation above we see that (3.5) is satisfied with $h(x \mid s) = \alpha\lambda[x^{\alpha+1} f(x/\lambda)]^{-1}$. In particular, the asymptotic upper bound for the normalized second moment is

$$\frac{1}{n^2} \int R(y) dm = \frac{1}{n^2} \lambda^{-2\alpha} \int_{1/\lambda}^{\infty} \frac{\alpha^2}{x^{2(\alpha+1)} f(x)} dx \sum_{i=1}^{n} \frac{1}{q_i} \prod_{j=1}^{i-1} \frac{1}{p_j},$$

with $q_n = 1$. It is straightforward to check that $\frac{1}{n^2} \sum_{i=1}^{n} \frac{1}{q_i} \prod_{j=1}^{i-1} \frac{1}{p_j}$ is minimized at

$$p_i = 1 - \frac{1}{n - i + 1}, \quad q_i = 1 - p_i,$$

with minimum equal to 1. The parameter λ can be chosen to control the factor

$$\lambda^{-2\alpha} \int_{1/\lambda}^{\infty} \frac{\alpha^2}{x^{2(\alpha+1)} f(x)} dx.$$

In some cases this can be minimized analytically.

Example 4.1. Consider a Pareto density of the form $f(x) = \alpha(1+x)^{-\alpha-1}$, $x > 0$. Then

$$\lambda^{-2\alpha} \int_{1/\lambda}^{\infty} \frac{\alpha^2}{x^{2(\alpha+1)} f(x)} dx = \lambda^{-2\alpha} \int_{1/\lambda}^{\infty} \alpha \Big(\frac{1+x}{x^2} \Big)^{\alpha+1} dx.$$

If $\alpha = 1$ this is minimized at $\lambda = \sqrt{3}$ with minimum $\frac{2+\sqrt{3}}{\sqrt{3}}$.

In the scaling mixture algorithm we assume $L_* > 0$. This rules out distributions whose slowly varying function tends to 0. However, this is not a severe problem. One way to avoid it is to slightly modify the previous algorithm. The *scaling mixture II* algorithm has, with $\lambda > 0$, $u \in (0,1)$, $\delta > 0$, and $a \in (0,1)$,

$$\begin{aligned} g_i(x \mid s) &= g(x) \\ &= (\lambda b)^{-1} f(x/\lambda b) I\{0 < x \leq \lambda b u\} \\ &\quad + \frac{1}{(1+\delta)\lambda b} \Big(\frac{x}{\lambda b} \Big)^{\frac{1}{1+\delta} - 1} f([x/\lambda b]^{\frac{1}{1+\delta}}) I\{x \geq \lambda b u^{1+\delta}\} + f(x) I\{x \leq 0\}, \\ g_n(x \mid s) &= g(x) I\{s \leq b - b(1-a)^{n-1}\} + f(x) I\{s > b - b(1-a)^{n-1}\}. \end{aligned}$$

The density g_i is based on the following sampling procedure. To generate a sample X from g_i, first generate a candidate X' from f. If $X' \leq 0$ put $X = X'$, if $0 < X' \leq u$, put $X = \lambda b X'$, and if $X' > u$ put $X = \lambda b (X')^{1+\delta}$.

Similar to the scaling mixture I algorithm it follows that, for $1 \leq i \leq n$,

$$\begin{aligned} &\frac{g_i(bx \mid bs)}{f(bx)} \overline{F}(b) \\ &\quad \to x^{\alpha+1} \frac{f(\frac{x}{\lambda})}{\alpha\lambda} I\{0 < x \leq \lambda u\} + \frac{x^{\frac{1}{1+\delta} + \alpha} f([x/\lambda]^{\frac{1}{1+\delta}})}{(1+\delta)\alpha\lambda^{\frac{1}{1+\delta}}} I\{x \geq \lambda u^{1+\delta}\} \end{aligned}$$

uniformly for $x \geq 1-s$ and s. Since

$$x^{\frac{1}{1+\delta}+\alpha} f([x/\lambda]^{\frac{1}{1+\delta}}) = \lambda^{-\frac{\alpha+1}{1+\delta}} x^{\alpha(1-\frac{1}{1+\delta})} L(x/\lambda),$$

is bounded from below for $x \geq 1-s$ (3.2) holds. Just as for the scaling mixture I algorithm (3.3) also holds. (3.5) hold with

$$h_i(y \mid s) = \frac{\alpha^2\lambda}{y^{2\alpha+2} f(y/\lambda)} + \frac{(1+\delta)\alpha^2\lambda^{\frac{1}{1+\delta}}}{y^{2\alpha+\frac{1}{1+\delta}+1} f([y/\lambda]^{\frac{1}{1+\delta}})}.$$

The asymptotic upper bound for the normalized second moment is hence

$$\int R(y)dm = \sum_{i=1}^{n} \frac{1}{q_i} \prod_{j=1}^{i-1} \frac{1}{p_j} \times \Big(\int_1^{\lambda u} \frac{\alpha^2\lambda}{x^{2\alpha+2} f(x/\lambda)} dx + \int_{\lambda u^{1+\delta}}^{\infty} \frac{(1+\delta)\alpha^2\lambda^{\frac{1}{1+\delta}}}{x^{2\alpha+\frac{1}{1+\delta}+1} f([x/\lambda]^{\frac{1}{1+\delta}})} dx \Big). \quad (4.4)$$

with $q_n = 1$. As above $\frac{1}{n^2}\sum_{i=1}^{n} \frac{1}{q_i} \prod_{j=1}^{i-1} \frac{1}{p_j}$ is minimized at

$$p_i = 1 - \frac{1}{n-i+1}, \quad q_i = 1 - p_i,$$

with minimum equal to 1. The remaining parameters λ and u can be chosen to control the integrals in (4.4).

5. Achieving asymptotically optimal relative error

In the previous section we observed that the conditional mixture algorithm and the GPD mixture algorithm can be designed to have almost asymptotically optimal relative error. A small asymptotic relative error is obtained by choosing the parameter a close to 1. In this section we prove that these algorithms have asymptotically optimal relative error. This is accomplished by letting the parameter a depend on the threshold b in such a way that $a \to 1$ slowly as $b \to \infty$. For simplicity, we assume that $X_1 > 0$ throughout this section.

Theorem 5.1. *Let ν_n^b be the measure defined by the conditional mixture algorithm. Let $p_i = \frac{n-i}{n-i+1}$, $q_i = 1 - p_i$, and assume that $1 - a = 1 - a_b \sim \mathcal{O}(b^{-\frac{1}{2(n-1)}+\delta})$ for some $0 < \delta < \frac{1}{2(n-1)}$. Then, the conditional mixture algorithm has asymptotically optimal relative error for computing p_b. That is,*

$$\lim_{b\to\infty} \frac{E\hat{p}_b^2}{p_b^2} = 1.$$

Remark 5.2. In Theorem 5.1 the conditional mixture algorithm can be replaced by the GPD mixture algorithm.

Proof. First rewrite the normalized second moment as in (3.1):

$$\frac{E\hat{p}_b^2}{p_b^2} = \frac{1}{n^2} \int\limits_{s_n>1} \frac{1}{\overline{F}(b)} \frac{d\mu_n}{d\nu_n^b}(by) m_b(dy) = \frac{1}{n^2} \int\limits_{s_n>1} R_b(y) m_b(dy).$$

Fix $a_0 \in (0,1)$. Define the sets

$$\begin{aligned}
A_i &= \{y \in \mathbf{R}^n : y_i > 1 - s_{i-1}, y_j \leq 1 - s_{j-1}, j < i\},\\
B_i &= \{y \in \mathbf{R}^n : y_i \leq a_0(1 - s_{i-1})\},\\
C_i &= \{y \in \mathbf{R}^n : a_0(1 - s_{i-1}) < y_i \leq a(1 - s_{i-1})\},\\
D_i &= \{y \in \mathbf{R}^n : a(1 - s_{i-1}) < y_i \leq 1 - s_{i-1}\}.
\end{aligned}$$

Then $\{s_n > 1\} \subset \cup_{i=1}^n A_i$ and each A_i can be written as the disjoint union of the 3^{i-1} sets of the form

$$I_1 \cap I_2 \cap \cdots \cap I_{i-1} \cap A_i, \tag{5.1}$$

where each I_j is either B_j, C_j, or D_j. Each intersection (5.1) is of one of the types below.

(i) $I_j = B_j$ for each $j = 1, \dots, i-1$.
(ii) among the sets $I_1, \dots, I_{i-1}$ there is at least one j for which $I_j = C_j$ and no j with $I_j = D_j$.
(iii) among the sets $I_1, \dots, I_{i-1}$ there is at least one j for which $I_j = D_j$.

Next we treat the integrals

$$\frac{1}{n^2} \int_{I_1 \cap \cdots \cap I_{i-1} \cap A_i} R_b(y) m_b(dy),$$

separately. The intersection belongs to one of the three types.

Type (i): Consider $y \in B_1 \cap \cdots \cap B_{i-1} \cap A_i$. Then $s_{i-1} \leq 1 - (1 - a_0)^{i-1}$ and

$$R_b(y) \leq \prod_{j=1}^{i-1} \frac{1}{p_j} \times \frac{1}{p_i \overline{F}(b) + q_i \frac{\overline{F}(b)}{\overline{F}(ba(1-s_{i-1}))}} \leq \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \frac{\overline{F}(ba(1 - s_{i-1}))}{\overline{F}(b)}.$$

Fix arbitrary $\varepsilon > 0$. Then, for b sufficiently large, $a_b > 1 - \varepsilon$ and the expression in the last display is bounded above by

$$\prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \frac{\overline{F}(b(1-\varepsilon)(1 - s_{i-1}))}{\overline{F}(b)} =: R_b^*(y).$$

It follows that $R_b^*(y) \to \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} (1-\varepsilon)^{-\alpha} (1 - s_{i-1})^{-\alpha}$ uniformly on $B_1 \cap \cdots \cap B_{i-1} \cap A_i$ and then it follows by the arguments in the proof of Theorem 3.2 that

$$\limsup_{b \to \infty} \frac{1}{n^2} \int_{B_1 \cap \cdots \cap B_{i-1} \cap A_i} R_b(y) m_b(dy) \leq \frac{1}{n^2} \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} (1-\varepsilon)^{-\alpha}.$$

Since $\varepsilon > 0$ was arbitrary we can let $\varepsilon \to 0$ to get

$$\limsup_{b \to \infty} \frac{1}{n^2} \int_{B_1 \cap \cdots \cap B_{i-1} \cap A_i} R_b(y) m_b(dy) \leq \frac{1}{n^2} \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i}.$$

Type (ii): For $y \in I_1 \cap I_2 \cap \cdots \cap I_{i-1} \cap A_i$ it holds that $s_{i-1} \le 1-(1-a)^{i-1}$. Proceeding as in the Type (i) case, for $\varepsilon > 0$ and b sufficiently large,

$$\begin{aligned} R_b(y) &\le \prod_{j=1}^{i-1} \frac{1}{p_j} \times \frac{1}{p_i \overline{F}(b) + q_i \frac{\overline{F}(b)}{\overline{F}(ba(1-s_{i-1}))}} \\ &\le \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \frac{\overline{F}(ba(1-s_{i-1}))}{\overline{F}(b)} \\ &\le \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \frac{\overline{F}(b(1-\varepsilon)(1-a)^{i-1})}{\overline{F}(b)}. \end{aligned}$$

It follows that

$$\begin{aligned} &\frac{1}{n^2} \int\limits_{I_1 \cap \cdots \cap I_{i-1} \cap A_i} R_b(y) m_b(dy) \\ &\quad \le \frac{1}{n^2} \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \frac{\overline{F}(b(1-\varepsilon)(1-a)^{i-1})}{\overline{F}(b)} m_b(I_1 \cap \cdots \cap I_{i-1} \cap A_i). \qquad (5.2) \end{aligned}$$

Let k be the first index between 1 and $i-1$ such that $I_k = C_k$. Then $I_l = B_l$ for each $1 \le l \le k-1$, and $s_{k-1} \le 1-(1-a_0)^{k-1}$, whereas $s_{i-1} \le 1-(1-a)^{i-1}$,

$$\begin{aligned} m_b(I_1 \cap \cdots \cap I_{i-1} \cap A_i) &\le \frac{P(Y_k > ba_0(1-s_{k-1}), Y_i > b(1-s_{i-1}))}{\overline{F}(b)} \\ &\le \frac{P(Y_k > ba_0(1-a_0)^{k-1}, Y_i > b(1-a)^{i-1})}{\overline{F}(b)} \\ &= \frac{\overline{F}(ba_0(1-a_0)^{k-1})\overline{F}(b(1-a)^{i-1})}{\overline{F}(b)} \\ &\le \frac{\overline{F}(ba_0(1-a_0)^{i-1})\overline{F}(b(1-a)^{i-1})}{\overline{F}(b)}. \qquad (5.3) \end{aligned}$$

Putting this into (5.2) yields the upper bound

$$\begin{aligned} &\frac{1}{n^2} \int\limits_{I_1 \cap \cdots \cap I_{i-1} \cap A_i} R_b(y) m_b(dy) \\ &\quad \le \frac{1}{n^2} \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \Big[\frac{\overline{F}(b(1-\varepsilon)(1-a)^{i-1})}{\overline{F}(b)} \Big]^2 \overline{F}(ba_0(1-a_0)^{i-1}) \\ &\quad \le \frac{1}{n^2} \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \Big[\frac{\overline{F}(b(1-\varepsilon)(1-a)^{n-1})}{\overline{F}(b)} \Big]^2 \overline{F}(ba_0(1-a_0)^{n-1}). \end{aligned}$$

This converges to 0 as $b \to \infty$ by the choice of $a = a_b$.

Type (iii): For $I_1 \cap I_2 \cap \cdots \cap I_{i-1} \cap A_i$ of type (iii) we let j denote the first index for which $I_j = D_j$. Suppose first that $I_k = B_k$ for each $k = 1, \dots, j-1$. Then, $s_{j-1} \le 1-(1-a_0)^{j-1}$ and, for arbitrary $\varepsilon > 0$ and b sufficiently large,

$$R_b(y) \le \prod_{k=1}^{j-1} \frac{1}{p_k} \frac{1}{q_j} \frac{\overline{F}(b(1-\varepsilon)(1-a_0)^{j-1})}{\overline{F}(b)}, \qquad (5.4)$$

which is bounded in b. In addition,

$$m_b(B_1 \cap \cdots \cap B_{j-1} \cap D_j) \le \frac{P(Y_j \in D_j, S_{j-1} \le 1-(1-a_0)^{j-1})}{\overline{F}(b)}$$

$$\le \int_{B_1 \cap \cdots \cap B_{j-1}} \frac{\overline{F}(ba(1-s_{j-1})) - \overline{F}(b(1-s_{j-1}))}{\overline{F}(b)} \mu_n(dy) \to 0, \tag{5.5}$$

as $b \to \infty$, by the bounded convergence theorem. Combining (5.4) and (5.5) we see that

$$\limsup_{b\to\infty} \int_{B_1\cap\cdots\cap B_{j-1}} R_b dm_b = 0.$$

Finally, suppose $I_k = C_k$ for some $k = 1, \dots, j-1$. Then, $s_{j-1} \le 1-(1-a)^{j-1}$ and, for arbitrary $\varepsilon > 0$ and b sufficiently large,

$$R_b(y) \le \prod_{k=1}^{j-1} \frac{1}{p_k}\frac{1}{q_j}\frac{\overline{F}(b(1-\varepsilon)(1-a)^{j-1})}{\overline{F}(b)}. \tag{5.6}$$

In addition, just as in (5.3),

$$m_b(I_1 \cap \cdots \cap I_{j-1} \cap D_j) \le \frac{\overline{F}(ba_0(1-a_0)^{j-1})\overline{F}(b(1-\varepsilon)(1-a)^{j-1})}{\overline{F}(b)}. \tag{5.7}$$

Combining (5.6) and (5.7) we see that

$$\limsup_{b\to\infty} \int_{B_1\cap\cdots\cap B_{j-1}} R_b dm_b$$

$$\le \limsup_{b\to\infty} \prod_{k=1}^{j-1} \frac{1}{p_k}\frac{1}{q_j}\Big[\frac{\overline{F}(b(1-\varepsilon)(1-a)^{j-1})}{\overline{F}(b)}\Big]^2 \overline{F}(ba_0(1-a_0)^{j-1}) = 0,$$

by the choice of $a = a_b$. □

6. Numerical illustrations

In this section we examine the performance of the scaling mixture algorithm, referred to as the SM algorithm. We perform a preliminary test using Pareto-distributed positive random variables and compare the algorithm with the conditional mixture algorithm in Dupuis et al. (2007), which we refer to as the DLW algorithm, and the conditional Monte Carlo algorithm in Asmussen and Kroese (2006). For comparision, we first consider the same setting as in Dupuis et al. (2007), Table IV, pp. 18. The so-called true value in Table 1 was obtained from the same table. Each estimate was calculated using $N = 10^4$ samples of S_n. This estimation was repeated 100 times and the mean estimate, the mean standard error and the mean calculation time were calculated. The parameter a in the DLW algorithm was chosen equal to 0.999 and the parameter λ in the scaling mixture algorithm was chosen equal to 1 in Table 1 and equal to $\sqrt{3}$, the optimal value, in Table 2.

The standard Monte Carlo estimation is inferior to both importance sampling algorithms. The conditional Monte Carlo algorithm performs best for most probabilites in this study.

TABLE 1. Simulations of $P(S_n > b)$, where $S_n = \sum_{i=1}^{n} X_i$ and $P(X_1 > x) = (1+x)^{-1/2}$, $a = 0.999$ and $\lambda = 1$. $N = 10^4$ samples were used for each estimation, repeated 100 times..

n	b	True value	MIS	DLW	CMC	MC	
5	5e+05	0.007071	0.0070744 (7.26e-05) [0.816]	0.0070714 (6.10e-06) [0.799]	0.00707034 (4.89e-06) [0.731]	0.0069960 (4.88e-05) [0.685]	Avg. est. (A. std. err.) [A. time (s)]
	5e+11	7.0711e-06	7.0776e-06 (7.53e-08) [1.005]	7.0710e-06 (1.86e-09) [0.990]	7.0711e-06 (2.71e-11) [0.908]	1.8000e-05 (1.56e-05) [0.840]	
15	5e+05	0.02121	0.021188 (2.07e-04) [1.224]	0.021215 (4.15e-05) [1.219]	0.021210 (2.72e-05) [1.092]	0.021724 (2.05e-03) [1.006]	
	5e+11	2.1213e-05	2.1224e-05 (2.25e-07) [1.450]	2.1214e-05 (5.82e-09) [1.456]	2.1213e-05 (3.09e-10) [1.283]	1.800e-05 (1.80e-05) [1.179]	
25	5e+05	0.035339	0.035330 (3.32e-04) [1.712]	0.035348 (9.06e-05) [1.729]	0.035347 (5.89e-05) [1.478]	0.035462 (2.61e-03) [1.366]	
	5e+11	3.5355e-05	3.5338e-05 (3.77e-07) [1.993]	3.5355e-05 (1.04e-09) [2.016]	3.5355e-05 (1.32e-09) [1.689]	3.8000e-05 (3.68e-05) [1.559]	

TABLE 2. Simulations of $P(S_n > b)$, where $S_n = \sum_{i=1}^{n} X_i$, $P(X_1 > x) = (1+x)^{-1}$, $a = 0.999$ and $\lambda = \sqrt{3}$. $N = 10^4$ samples were used for each estimation, repeated 100 times.

n	b	True value	MIS	DLW	CMC	MC	
5	5e+05	1.0001e-05	1.0020e-05 (1.07e-07) [0.429]	1.0001e-05 (2.78e-09) [0.415]	1.0001e-05 (2.58e-10) [0.433]	6.000e-06 (6.00e-6) [0.346]	Avg. est. (std. err.) [time (s)]
	5e+11	1.0000e-13	9.9996e-12 (1.07e-13) [0.433]	9.9999e-12 (2.79e-15) [0.418]	1.0000e-13 (8.59e-22) [0.430]	0 (0) [0.352]	
15	5e+05	3.0010e-05	3.0004e-05 (3.21e-07) [0.491]	3.0011e-05 (1.12e-08) [0.445]	3.0010e-05 (1.74e-09) [0.437]	3.0000e-05 (2.71e-05) [0.375]	
	5e+11	3.0000e-11	2.9990e-11 (3.22e-13) [0.490]	3.0000e-11 (9.06e-15) [0.445]	3.0000e-11 (1.75e-20) [0.431]	0 (0) [0.365]	
25	5e+05	5.0029e-05	5.0098e-05 (5.37e-07) [0.561]	5.00274e-05 (1.90e-08) [0.485]	5.00290e-05 (4.10e-09) [0.432]	3.7000e-05 (3.34e-05) [0.386]	
	5e+11	5.0000e-11	4.9970e-11 (5.38e-13) [0.556]	4.9998e-11 (1.65e-14) [0.479]	5.0000e-11 (1.54e-20) [0.439]	0 (0) [0.382]	

REFERENCES

Asmussen, S. and Binswanger, K., 1997. Ruin probability simulation for subexponential claims. ASTIN Bull. 27, 297-318.

Asmussen, S. and Glynn, P.W., 2007. Stochastic Simulation: Algorithms and Analysis. Springer-Verlag, New York.

Asmussen, S. and Kroese, D.P., 2006. Improved algorithms for rare event simulation with heavy tails. Adv. Appl. Probab. 38, 545-558.

Bassamboo, A., Juneja, S. and Zeevi, A., 2007. On the efficiency loss of state-independent importance sampling in the presence of heavy-tails. Oper. Res. Lett. 35, 251-260.

Blanchet, J. and Li, C., 2008. Efficient rare event simulation for heavy-tailed compound sums. Preprint, Columbia University.

Blanchet, J. and Liu, J.C., 2008. State-dependent importance sampling for regularly varying random walks. Adv. Appl. Probab. 40, 1104-1128.

Dupuis, P., Leder, K. and Wang, H., 2007. Importance sampling for sums of random variables with regularly varying tails. ACM Trans. Model. Comput. Simul. 17(3).

Juneja, S. and Shahabuddin, P., 2002. Simulating heavy-tailed processes using delayed hazard rate twisting. ACM Trans. Model. Comput. Simul. 12, 94-118

Resnick, S.I., 1987. Extreme Values, Regular Variation, and Point Processes, Springer-Verlag, New York.

(H. Hult) Department of Mathematics, KTH, 100 44 Stockholm, Sweden
E-mail address: `hult@kth.se`

(J. Svensson) Department of Mathematics, KTH, 100 44 Stockholm, Sweden
E-mail address: `jenssve@kth.se`